7.3 Статистична обробка даних

При обробці результатів виникає необхідність оцінювання характеристик випадкової величини (перед вивченням цього розділу рекомендовано ознайомитись з розділом 1.5.2.1 підручника).

Як оцінка $\overline{X}$ невідомого математичного сподівання m_X випадкової величини X використовується середнє арифметичне результатів N незалежних випробувань

$$\overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} x_i}{N},$$

а для оцінки дисперсії D_x при достатньо великій кількості експериментальних даних ($N \geq 30$) співвідношення

$$\overline{D}_x = \overline{\sigma}_X^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1}.$$

При припущенні нормального закону розподілу величини X можна показати, що величина

$$T = \frac{\overline{x} - m_X}{\sigma_X / \sqrt{N}}$$

має *t*-розподіл Стьюдента з $k{=}N-1$ ступенями свободи (ступінь свободи в статистиці визначається як різниця між кількістю дослідів та кількістю коефіцієнтів моделі, які можна обчислити за результатами цих експериментів незалежно один від одного, наприклад, у нормального розподілу два параметри, у Пуассонівського – один і т.д.) Звідси можна визначити надійний інтервал для значення $\boldsymbol{X}$: за відомими значеннями надійної імовірності P з таблиці 7.3 знаходимо ε, звідки

$$\Delta = \varepsilon * \frac{D_X}{\sqrt{N}}.$$

Таким чином, якщо випадкова величина x розподілена за нормальним законом з математичним сподіванням m_x і дисперсією D_X, то значення x знаходиться в інтервалі $(m_x - \Delta, m_x + \Delta)$ з імовірністю P.

Для оцінювання виду закону розподілу при довільному законі розподілу найширше застосування мають критерії Колмогорова та Пірсона (критерій Стьюдента застосовується тільки при нормальному законі розподілу) , що дозволяють на основі порівняння емпіричної функції розподілу $f_x^*(x)$, одержаної у вигляді гістограми в результаті обробки експериментальних даних, з гіпотетичною $f_x(x)$, яка відповідає запропонованій гіпотезі, зробити висновки про їх збігання чи незбігання при рівні значущості α, який визначається як імовірність того, що буде відхилена достовірна гіпотеза.

Таблиця 7.3 – Значення ε для інтервалу $-\varepsilon < t < \varepsilon$, де величина t має розподіл Стьюдента залежно від надійної імовірності P і числа ступенів свободи k

k	P =0,90	P =0,95	P =0,99
1	6,310	12,71	63,7
2	2,920	4,30	9,92
3	2,350	3,18	5,84
4	2,130	2,77	4,60
5	2,020	2,57	4,03
6	1,943	2,45	3,71
7	1,895	2,36	3,50
8	1,860	2,31	3,36
9	1,833	2,26	3,25
10	1,812	2,23	3,17
11	1,796	2,20	3,11
12	1,782	2,18	3,06
13	1,771	2,16	3,01
14	1,761	2,14	1,98
15	1,753	2,13	2,95
16	1,746	2,12	2,92
17	1,740	2,11	2,90
18	1,734	2,10	2,86
19	1,729	2,09	2,86
20	1,725	2,08	2,84
22	1,717	2,07	2,82
24	1,711	2,06	2,80
26	1,706	2,06	2,78
28	1,701	2,05	2,76
30	1,697	2,04	2,75
40	1,684	2,02	2,70
60	1,671	2,00	2,66
120	1,658	1,98	2,62
240	1,645	1,96	2,58

В критерії Колмогорова мірою є величина

$$\lambda = \left| f_X(x) - f_X^*(x) \right|_{\max} \sqrt{N},$$

яку порівнюють з критичним значенням, заданим з таблиці 7.4.

Таблиця 7.4 – Критичні значення λ залежно від рівня значущості

α	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,001	0,001
$\lambda_{кр}$	0,828	0,895	0.974	1,073	1,224	1,358	1,520	1,627	1,950

При $\lambda < \lambda_{кр}$ гіпотеза про збігання $f_X(x)$ и $f_X^*(x)$ приймається.

В критерії Пірсона обчислюється величина

$$\chi^2 = \sum_{i=0}^{k} \frac{\left[f_X(x_i) - f_X^*(x_i) \right]^2}{f_X(x_i)},$$

де k – кількість розрядів гістограми (дискретних значень $f_X(x_i)$).

З таблиці 7.5 визначають критичне значення χ^2, зважаючи на α і число ступенів свободи

$$r = k - l - 1,$$

де l – число параметрів, що їх містить в собі закон розподілу (як вже відмічалося :для нормального =2, Пуассонівського l=1 і т.д.).

При $\chi^2 \langle \chi^2_{кр}$ гіпотеза приймається.

Якщо порівнюють аналітично одержані закони розподілу ймовірностей, то мірою їх близькості слугує значення середньої квадратичної похибки.

Для оцінювання взаємозалежності випадкових величин, між якими існує стохастичний зв’язок, використовується коефіцієнт кореляції (correlation coefficient) (n – обсяг вибірки)

$$r_{xy} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_x)(y_i - m_Y)}{\sigma_x \sigma_y}.$$

При визначенні взаємозалежності значень випадкових величин в різні моменти часу коефіцієнт кореляції оцінюється за формулою

$$r_{X(\tau)} = \frac{\frac{1}{n-m-1}\sum_{i=1}^{n-m}\left[x(t_i) - m_x\right]\left[x(t_i + \tau) - m_X\right]}{D_X},$$

де $x(t_i)$ – значення випадкової величини X в момент часу t_i, а $x(t_i+\tau)$ – в момент часу, який відрізняється від t_i на інтервал τ. Таким чином, $x(t_i)=x_i$, $x(t_i+\tau)=x_j$, τ – інтервал часу між *i* та *j* значеннями x, $i-j=m$.

Таблиця 7.5 – Критичні точки розподілу x – випадкова величина, яка розподілена за законом χ^2 з ступенями свободи *k*

Число ступенів свободи k	$\alpha =$ **0,01**	α **=0; 0,025**	α = **0,05**	α = **0,95**	α = **0,975**	α = **0,99**
1	6,6	6,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81

Інтервал кореляції визначається як відрізок часу, за який кореляційна функція зменшується на 95%.

Обчислення коефіцієнту кореляції (нормованої кореляційної функції) та кореляційної функції за відомими масивами даних x та y за допомогою наведених формул не викликає труднощів, а апроксимація вигляду кореляційної функції типовими кореляційними функціями (таблиця 7.6) може здійснюватися за методом найменших квадратів.

Таблиця 7.6 – Типові кореляційні функції

Вигляд	Параметри
$R_x(\tau)=\sigma_X^2(1-\alpha\mid\tau\mid)$, $\tau<1/\alpha$	$\alpha=(\sigma_X^2-R_x(\tau^*))/\sigma_X^2\tau$, $R_x(\tau^*)$ – відоме значення кореляційної функції
$R_x(\tau)=\sigma_X^2 e^{-\alpha\|\tau\|}$	$\alpha=\frac{1}{\tau^*}\ln\frac{\sigma_x^2}{R(\tau^*)}$
$R_x(\tau)=\sigma_X^2 e^{-\alpha^2\tau^2}$	$\alpha=\frac{1}{\tau^*}\sqrt{\ln\frac{\sigma_x^2}{R_x(\tau^*)}}$
$R_x(\tau)=\sigma_X^2 e^{-\alpha\|\tau\|}(1+\alpha\mid\tau\mid)$	$\alpha\approx 4{,}5/\tau_k^{\max}$
$R_x(\tau)=\sigma_X^2 e^{-\alpha\|\tau\|}\cos\beta\tau$	$\alpha=\frac{\ln}{\tau_2^*}\cdot\frac{\sigma_x^2\cos\frac{\pi\tau_2^*}{2\tau_1^*}}{R_x(\tau_2^*)}$, $\beta=\pi/2\tau_1^*$ при двох відомих значеннях кореляційної функції $R_x(\tau)$, причому $R_x(\tau_1^*)=0$.